Beyond Dysplasia: Uncovering Structure in Oral Potentially Malignant Diseases with Unsupervised Contrastive Learning

Kayla Caughlin[1], Rodrigo Cuenca Martinez[2], Gabriel P. Tortorelli[2], Yi-Shing L. Cheng[3], Rashmi Hegde[3], Celeste Abraham[3], Jacqueline M. Plemons[3], Ying S. Wang[3], Victoria Woo[3], Javier A. Jo[2], and Carlos Busso[1]

***Abstract*— Automated cancer diagnosis research often focuses on a binary task - recognize dysplasia and cancer from other lesions. However, other clinical conditions have estimated malignant transformation rates. Grouping these oral potentially malignant diseases with benign conditions may not be ideal. While automated cancer diagnosis potential has been shown in multi-spectral autofluorescence images, the existence of a more detailed structure has not been investigated. Training multi-class models is often avoided on small data sets due to the low number of samples per class, but standard clustering algorithms like k-means lack supervision to manage high inter-patient variability. We propose an unsupervised contrastive clustering approach, where a small neural network is trained to group normal images together, away from lesion images. We expect our method to reduce inter-patient variability and amplify the existing structure among different types of oral lesions. Our results indicate that a structure based on risk level for malignant transformation does exist, generating a cluster composed of over 75% high-risk lesions and out-performing the k-means baseline of 40%.**

***Clinical relevance*— Instead of focusing on the identification of dysplasia and squamous cell carcinoma from other lesions in a binary task, our investigational analysis shows that more detailed structure exists in multi-spectral autofluorescence images, with applications in the management of oral potentially malignant diseases.**

I. INTRODUCTION

Several recent publications have described advancements in the equipment and algorithms used to automatically and non-invasively detect oral cancer [1]–[4]. However, the focus of the classification algorithms is typically a binary task to distinguish benign from malignant lesions. As an alternative to a biopsy, the binary classification task is appropriate. At the time of the biopsy, the lesion has already been identified as suspicious and in need of histological investigation for a more conclusive diagnosis. However, in non-invasive systems, images may be taken from a wider variety of lesions and of normal tissue. In *multi-spectral autofluorescence lifetime imaging* (maFLIM), data collection may include a variety of clinical diagnoses such as morsicatio, lichen planus, leukoplakia, geographic tongue, and fibroma. While automated maFLIM diagnosis models have been trained to separate dysplasia and cancer from other lesions, *oral potentially malignant diseases* (OPMD) such as leukoplakias are not benign and have defined malignant transformation rates. For example, leukoplakia, erythroplakia, and *proliferative verrucous leukoplakia* (PVL) have estimated malignant transformation rates of 9.5%, 33.1%, and 49.5%, respectively [5]. While Iocca et al. [5] listed lichen planus with the OPMDs, the estimated rate of malignant transformation is only 1.4% and the inclusion of lichen planus as an OPMD has been debated.

Based on the differing malignant transformation rates, a risk-stratification approach better reflects the differences in clinical diagnoses that have previously been lumped into a single non-dysplastic, non-cancerous group. A non-invasive fluorescence based imaging system may produce images with more granular information than the binary framework for identification of dysplasia and squamous cell carcinoma. We are interested in understanding if maFLIM images contain information for risk stratification of oral lesions for use in a general dentistry setting, where the device would be used to find patients most at risk for the development of cancer. Such a device would allow specialists to focus on patients most in need and decrease time to treatment.

Implementing a risk-stratification model with standard supervised training, however, may be impractical due to data scarcity and inter-patient variability. When the number of classes is increased to account for the different malignant transformation rates, the amount of training data for each class decreases and performance is likely to drop for small data sets. Furthermore, separating lesions by OPMD group with a strict label will be noisy, as each group will contain lesions that eventually develop into cancer at different rates, with no current ability to differentiate between these outcomes. We argue that unsupervised methods are interesting to avoid splintering the data set into many (noisy) classes. However, unsupervised clustering like K-means may not show the expected structure due to inter-patient variability masking the desired signal. To address these challenges, this work investigates potential data structures in maFLIM images that reflect natural groupings using a contrastive clustering method that leverages the diversity of images that can be collected using a non-invasive imaging system.

Rather than enforcing invariance between a fluorescence image and its augmentation, we focus on using the readily available groups of normal images and lesion images. We

*This work was supported by NIH R01:5R01CA218739

[1]The Department of Electrical and Computer Engineering, The University of Texas at Dallas, Richardson, TX 75080, USA. email@school.edu

[2]The School of Electrical and Computer Engineering, The University of Oklahoma, Norman, OK, 73019, USA.

[3]Texas A&M School of Dentistry, Dallas, TX 75246, USA.

train a small neural network model using a contrastive loss that learns to group normal images together and push away lesion images. We expect that by grouping the normal images, the model will separate patient-specific information from the disease process signal and uncover the natural groupings in the data. By avoiding any class labels for the lesion images, we do not have the problem of small sample sizes in a large number of classes. We compare our method to the k-means algorithm on the image features and find that the proposed contrastive clustering increases the density of higher-risk lesions such as leukoplakia in a single cluster.

II. IMAGING SYSTEM

Images were collected using a dual-excitation (375nm and 445nm), frequency domain maFLIM imaging system as detailed by Serafino et al. [1], with preliminary data reported by Cuenca et al. [2]. The system captures autofluorescence from endogenous fluorophores previously linked to malignant transformation. The four emission bands were selected to target collagen, *nicotinamide adenine dinucleotide* (NADH), reduced *flavin adenine dinucleotide* (FAD), and porphyrins. To understand the theoretic background related to the selection of endogenous fluorophores, we refer the reader to previous studies [6]–[14]. The collected autofluorescence signal is transformed into the frequency domain on a *field programmable gate array* (FPGA) before storage in the computer. From the magnitude of the frequency representation, we calculate normalized intensity features of the signal at each spatial location as described in Eqs. 1 and 2. The image size is 160x160 pixels.

There are four emission bands from the 345nm laser and three emission bands from the 445nm laser, resulting in 7 absolute intensity features. We use two normalization strategies, normalizing the absolute intensity features across the emission bands or across the excitation lasers. We denote the features normalized across the emission bands as *inter*-normalization. We calculate the inter-normalized intensity for the 405nm emission band as below:

$$I_{ex,em_{405}} = \frac{I_{ex,em_{405}}}{I_{ex,em_{405}} + I_{ex,em_{475}} + I_{ex,em_{550}} + I_{ex,em_{647}}} \quad (1)$$

where I represents the normalized intensity, ex represents the excitation wavelength, and em represents the emission wavelength. The inter-normalized features for the other emission bands are calculated using the same process. To avoid redundancy, we do not use the inter-normalized features from the 647nm emission band, resulting in 5 inter-normalized intensity features. Similarly, we denote features normalized across excitation bands as *intra*-normalized intensity features. For a given emission band, we calculate the features as below:

$$I_{ex_a,em} = \frac{I_{ex_a,em}}{I_{ex_a,em} + I_{ex_b,em}} \quad (2)$$

We also include phase and modulation features for a total of 50 features. We note that the signal representation for each pixel with the frequency domain system differs from our previously reported work [4], [15]–[17]. The previous data representation was a time series at each pixel, while the current representation is a complex number at each pixel for each excitation-emission pair.

III. DATA

The data in this analysis was collected at Texas A&M College of Dentistry (IRB2018-1251). This is an ongoing effort where the number of patients included in the data has increased since the start of the analysis. In this study, we used the first 88 patients. The data collection protocol allows for imaging of both normal and lesion tissues. Typically, a lesion image is collected at the center of the lesion and a normal image is taken from the contralateral, normal side. In the case of bilateral involvement, the normal image is taken from the most visually normal appearing tissue. The data set consists of a wide variety of lesions, with a high proportion of lichen planus. Thus, we are interested in looking for structure in three groups based on malignant transformation rates. The first group is the risk-based cases, where we group leukoplakia, PVL, dysplasia, and *squamous cell carcinoma* (SCC) as a high-risk group. We would also include any erythroplakia or erythro-leukoplakia in the high-risk group if patients with these lesions are enrolled in the study in the future. The second group is the mid-risk cases consisting of lichen planus and related lesions such as lichenoid lesions. A recent article discusses the uncertainty of whether dysplasia exhibits lichen planus characteristics or if lichen planus is a precursor to dysplasia [18]. Due to the controversial placement of lichen planus as a potentially malignant disease, we place these in a separate gray-area grouping. While Iocca et al. show that lichen planus has a very low malignant transformation rate [5], whether or not lichen planus is pre-malignant has been debated. A 2016 position paper from the American Academy of Oral and Maxillofacial Pathology recommends that no histologically evident dysplasia be present for a true lichen planus diagnosis [19]. The 2003 modified WHO criterion also maintains that the presence of dysplasia would rule out a true lichen planus diagnosis [20]. The third group is the low-risk cases consisting of other lesions that are not expected to develop cancer, including fibromas, ulcers, fissure tongue, geographic tongue, and pigmentation.

IV. PROPOSED APPROACH

A. Unsupervised Contrastive Clustering

Contrastive learning methods are increasingly used in natural image processing applications and include both unsupervised and supervised versions. In general, the goal of contrastive learning methods is to generate a feature representation that is organized according to pre-defined constraints. For example, *simple contrastive learning of visual representations* (SimCLR) [21] trains a network such that the features of an image and its augmentation (a positive pair) are similar, while other images and their augmentations (negative pairs) are more distant in the feature space. *Bootstrap your*

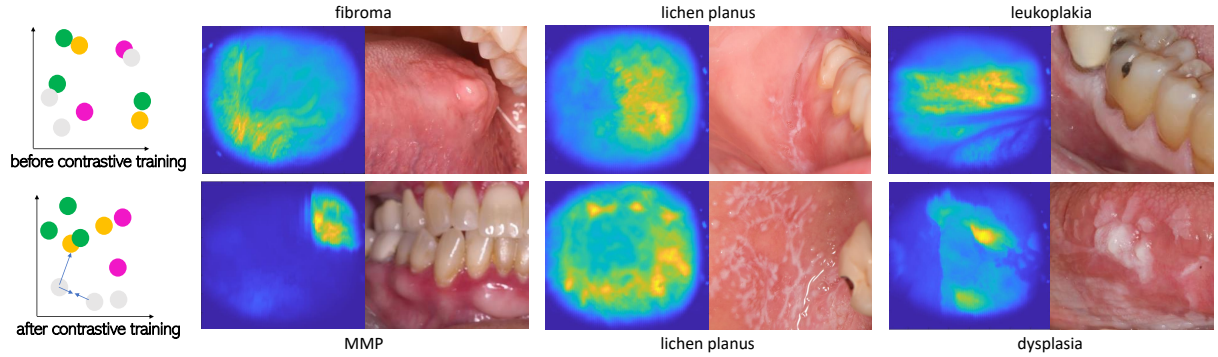

Fig. 1: Examples of lesions in our database and proposed method. Top left: normal images (gray data points) and lesion images (colored data points from different clinically-relevant groupings) do not show the desired structure. Bottom left: after contrastive training, the normal images group together away from lesion images, which show clinically relevant structure. Right: fibroma and *mucous membrane pemphigoid* (MMP) are examples of non-neoplastic (low risk) lesions, lichen planus is an example of the mid-risk group, and leukoplakia and dysplasia are examples of the high risk group.

own latent (BYOL) [22] and *simple siamese representation learning* (SimSiam) [23] are different implementations of contrastive learning using only positive pairs. As noted by the authors of BYOL, while multiple image augmentation methods exist, finding the best augmentation method for contrastive learning in other methods is not established [22]. Since the fluorescence lifetime imaging modality produces images based on the autofluorescence of cellular structures and metabolism and is typically trained on a pixel level, standard fluorescence lifetime image augmentation methods are not yet well-defined.

To improve the clustering results, we apply a contrastive clustering method to teach a small neural network to group normal pixels together, away from the lesion pixels. We do not impose a specific direction for the lesion pixels to move away, nor do we use any diagnoses for the lesions. We use a simple loss function that penalizes the model for normal samples that are far from the normal center by contributing positively to the loss function. The lesion to normal center distance contributes a negative term to the loss function, encouraging the lesion pixels to move away from the normal. With this approach, the network learns to minimize inter-patient variability among normal images and allows more clinically relevant structures to naturally appear in the lesion images.

B. Implementation

The network structure is a simple three-layer fully connected neural network with a dropout rate set at 0.5. The network takes the 50-dimensional feature vector as input and reduces the dimensionality to an 8-dimensional embedding output. We use the ADAM optimizer and train for 50 epochs. We create three clusters on the 8-dimensional output using the K-means algorithm with three centroids.

V. RESULTS AND ANALYSIS

The proposed approach was evaluated with the data described in Section III. Table I reports the results. The three clusters contain 18.87%, 18.18%, and 76.47% of high-risk lesions, respectively. This result indicates that most high-risk lesions are automatically grouped in one of the clusters created by our approach. As a baseline, we use the K-means clustering with three centroids on the original 50-dimensional feature vector. The feature level K-means results are shown in the center column of Table I. The three clusters contain 23.56%, 28.07%, and 38.89% of the leukoplakia or higher risk group. While cluster C does have a higher percentage of the high-risk group that indicates some structure, the density of the cluster is not as effective as our contrastive clustering strategy.

TABLE I: The proposed method increases density of high risk lesions in a single cluster. The middle column shows the percent of lesions at each risk level for the baseline k-means approach using image features. The right column shows the percent of lesions at each risk level for the proposed, contrastive approach.

	K-means Features			K-means Contrastive Emb.		
Cluster	Low	Mid	High	Low	Mid	High
A	47.06%	29.41%	23.53%	41.51%	39.62%	18.87%
B	31.58%	40.35%	28.07%	31.82%	50.00%	18.18%
C	27.78%	33.33%	38.89%	11.76%	11.76%	76.47%

The contrastive clustering improves the high-risk density of the top cluster from 38.89% to 76.47%, nearly doubling the high-risk density. We look at the samples included in cluster C (Table II). In addition to four lower-risk images in this high-risk cluster, one is a smokeless tobacco keratosis lesion. Since tobacco use is a known risk factor for oral cancer [24], placing this lesion in the high-risk group makes sense for risk stratification approaches. We also note that some ambiguity in the clinical diagnoses exists. For example, many patients recruited for the study with lichen planus had very well-managed lichen planus that had been previously treated at the time of imaging. These images are likely to be very similar to the low-risk or normal images, while other

TABLE II: Specific diagnoses in each cluster for the K-means contrastive Embedding.

Cluster A	Cluster B
Dysplasia (4)	Dysplasia (1)
Fibroma (3)	Fissure tongue (1)
Geographic tongue (2)	Hyperkeratosis (1)
Hyperkeratosis (3)	Leukoplakia (3)
Hyperplasia (1)	Lichen planus/lichenoid mucositis (11)
Hypersensitivity, plasma cell gingivitis (1)	Melanosis (1)
Leukoedema (1)	MMP (1)
Leukoplakia/PVL (5)	Trauma from chipped tooth (1)
Lichen planus/lichenoid mucositis (21)	Ulcer (2)
MMP (1)	
Morsicatio (3)	Cluster C
Non-specific, denture related erosion (1)	Dysplasia (6)
SCC (1)	Leukoplakia/PVL (5)
Ulcer (3)	Lichen planus (2)
Unsuspicious area that sloughs off (1)	Morsicatio (1)
Unusual - no clear diagnosis (1)	SCC (2)
Vascular lesion (1)	Smokeless Tobacco Keratosis (1)

images from patients with lichen planus had very severe lesions. These differences may explain why some lichen planus images clustered in cluster B, away from the cluster with a high portion of low-risk lesions.

In addition to the clear high-risk grouping in cluster C with the contrastive approach, the unsupervised approach has some diagnostic potential for a binary task despite not using labels. If we combine the two smaller clusters (clusters B and C) to form a "high risk positive" group and set the largest cluster (cluster A) as a "high risk negative" group, we get a sensitivity of 62.96% and a specificity of 66.15%. We combine the two smaller clusters because these clusters do not contain any normal images. Without prior knowledge, we would not know which of the two smaller clusters should correspond to the high-risk group since we do not train with labels.

VI. CONCLUSIONS

Our analysis indicates that more detailed information can be uncovered in multi-spectral autofluorescence images of oral lesions beyond a binary separation of dysplasia/SCC from other lesions. Using a fine-grained clinical decision support system, clinicians can optimize a patient's treatment by understanding the patient's risk level for developing oral cancer. With a non-invasive option to determine risk level, patients with low risk can avoid a specialist consultation and invasive biopsy, reducing wait times for patients with the most severe cases.

The proposed unsupervised contrastive clustering framework improves the clustering performance over a baseline using the K-means implemented with the original features. Our proposed approach increases the maximum high-risk cluster density from less than 40% to over 75%. We note that this performance may be increased by additional techniques like training with pseudo-labels or further training in a semi-supervised manner with a small set of labeled data. We plan to explore these directions in our future work. Our results are a proof-of-concept indicating structure in the features for different levels of oral lesions. We expect that the results will improve as more data from the ongoing research is collected. Finally, we are also interested in understanding the location-specific patterns that may be reflected in the response observed for different types of lesions [25].

REFERENCES

[1] M. Serafino, B. Applegate, and J. Jo, "Direct frequency domain fluorescence lifetime imaging using field programmable gate arrays for real time processing," *Review of Scientific Instruments*, vol. 91, no. 3, p. 033708, March 2020.

[2] R. Cuenca, M. J. Serafino, G. P. Tortorelli, R. C. Faram, K. Higgins, S. Khajotia, Y.-S. L. Cheng, J. M. Plemons, V. Woo, C. Busso, K. Caughlin, and J. Jo, "Dual-excitation multispectral autofluorescence lifetime endoscopy for clinical label-free metabolic imaging of oral lesions," in *Imaging, Therapeutics, and Advanced Technology in Head and Neck Surgery and Otolaryngology 2023*, vol. SPIE PC12354, San Francisco, CA, USA, January-February 2023, pp. 12 354–8.

[3] M. Marsden, B. Weyers, J. Bec, T. Sun, R. Gandour-Edwards, A. Birkeland, M. Abouyared, A. Bewley, D. Farwell, and L. Marcu, "Intraoperative margin assessment in oral and oropharyngeal cancer using label-free fluorescence lifetime imaging and machine learning," *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 3, pp. 857–868, March 2021.

[4] K. Caughlin, E. Duran-Sierra, S. Cheng, R. Cuenca, B. Ahmed, J. Ji, M. Martinez, M. Al-Khalil, H. Al-Enazi, Y.-S. L. Cheng, J. Wright, J. Jo, and C. Busso, "Aligning small datasets using domain adversarial learning: Applications in automated in vivo oral cancer diagnosis," *IEEE Journal of Biomedical and Health Informatics*, vol. 27, no. 1, pp. 457–468, January 2023.

[5] O. Iocca, T. Sollecito, F. Alawi, G. Weinstein, J. Newman, A. De Virgilio, P. Di Maio, G. Spriano, S. Pardiñas López, and R. Shanti, "Potentially malignant disorders of the oral cavity and oral dysplasia: A systematic review and meta-analysis of malignant transformation rate by subtype," *Head & neck*, vol. 42, no. 3, pp. 539–555, March 2020.

[6] I. Pavlova, M. Williams, A. El-Naggar, R. Richards-Kortum, and A. Gillenwater, "Understanding the biological basis of autofluorescence imaging for oral cancer detection: high-resolution fluorescence microscopy in viable tissue," *Clinical Cancer Research*, vol. 14, no. 8, pp. 2396–2404, April 2008.

[7] C. Gulledge and M. Dewhirst, "Tumor oxygenation: a matter of supply and demand," *Anticancer research*, vol. 16, no. 2, pp. 741–749, March-April 1996.

[8] R. Drezek, C. Brookner, I. Pavlova, I. Boiko, A. Malpica, R. Lotan, M. Follen, and R. Richards-Kortum, "Autofluorescence microscopy of fresh cervical-tissue sections reveals alterations in tissue biochemistry with dysplasia," *Photochemistry and photobiology*, vol. 73, no. 6, pp. 636–641, June 2001.

[9] N. Ramanujam, R. Richards-Kortum, S. Thomsen, A. Mahadevan-Jansen, M. Follen, and B. Chance, "Low temperature fluorescence imaging of freeze-trapped human cervical tissues," *Optics Express*, vol. 8, no. 6, pp. 335–343, March 2001.

[10] B. Chance, B. Schoener, R. Oshino, and F. I. abd Y. Nakase, "Oxidation-reduction ratio studies of mitochondria in freeze-trapped samples. NADH and flavoprotein fluorescence signals," *Journal of Biological Chemistry*, vol. 254, no. 11, pp. 4764–4771, June 1979.

[11] Z. Zhang, D. Blessington, H. Li, T. Busch, J. Glickson, Q. Luo, B. Chance, and G. Zheng, "Redox ratio of mitochondria as an indicator for the response of photodynamic therapy," *Journal of Biomedical Optics*, vol. 9, no. 4, pp. 772–778, July/August 2004.

[12] M. Müller, T. Valdez, I. Georgakoudi, V. Backman, C. Fuentes, S. Kabani, N. Laver, Z. Wang, C. Boone, R. Dasar, S. Shapshay, and M. Feld, "Spectroscopic detection and evaluation of morphologic and biochemical changes in early human oral carcinoma," *Cancer: Interdisciplinary International Journal of the American Cancer Society*, vol. 97, no. 7, pp. 1681–1692, April 2003.

[13] A. Shah, M. Beckler, A. Walsh, W. Jones, P. Pohlmann, and M. Skala, "Optical metabolic imaging of treatment response in human head and neck squamous cell carcinoma," *Plos One*, vol. 9, no. 3, p. e90746, March 2014.

[14] J. Jo, B. . Applegate, J. Park, S. Shrestha, P. Pande, I. Gimenez-Conti, and J. Brandon, "In vivo simultaneous morphological and biochemical optical imaging of oral epithelial cancer," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 10, pp. 2596–2599, October 2010.

[15] K. Caughlin, E. Duran-Sierra, S. Cheng, R. Cuenca, B. Ahmed, J. Ji, V. Yakovlev, M. Martinez, M. Al-Khalil, H. Al-Enazi, J. A. Jo, and C. Busso, "End-to-end neural network for feature extraction and cancer diagnosis of in vivo fluorescence lifetime images of oral lesions," in *International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2021)*, Guadalajara, Mexico, November 2021, pp. 3894–3897.

[16] P. Vasanthakumari, R. Romano, R. Rosa, A. Salvio, V. Yakovlev, C. Kurachi, and J. Jo, "Classification of skin-cancer lesions based on fluorescence lifetime imaging," in *SPIE Medical Imaging, 2020: Biomedical Applications in Molecular, Structural, and Functional Imaging*, vol. 11317, Houston, TX, USA, February 2020.

[17] E. Duran-Sierra, S. Cheng, R. Cuenca, B. Ahmed, J. Ji, V. Yakovlev, M. Martinez, M. Al-Khalil, H. Al-Enazi, Y. Cheng, J. Wright, C. Busso, and J. Jo, "Machine-learning assisted discrimination of precancerous and cancerous from healthy oral tissue based on multispectral autofluorescence lifetime imaging endoscopy," *Cancers*, vol. 13, no. 19, pp. 1–16, September 2021.

[18] J.-W. Li, K. Li, B. Chan, C. McGrath, and L.-W. Zheng, "Rate of malignant transformation differs based on diagnostic criteria for oral lichenoid conditions: A systematic review and meta-analysis of 24,277 patients," *Cancers*, vol. 15, no. 9, p. 2537, April 2023.

[19] Y.-S. L. Cheng, A. Gould, Z. Kurago, J. Fantasia, and S. Muller, "Diagnosis of oral lichen planus: a position paper of the american academy of oral and maxillofacial pathology," *Oral Surgery, Oral Medicine, Oral Pathology and Oral Radiology*, vol. 122, no. 3, pp. 332–354, September 2016.

[20] E. H. Van Der Meij and I. Van Der Waal, "Lack of clinicopathologic correlation in the diagnosis of oral lichen planus based on the presently available diagnostic criteria and suggestions for modifications," *Journal of oral pathology & medicine*, vol. 32, no. 9, pp. 507–512, October 2003.

[21] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A simple framework for contrastive learning of visual representations," in *International Conference on Machine Learning (ICML 2020)*, ser. Proceedings on Machine Learning, H. Daumé III and A. Singh, Eds. Vienna, Austria: PMLR, July 2020, vol. 119, pp. 1597–1607.

[22] J.-B. Grill, F. Strub, F. Altché, C. Tallec, P. Richemond, E. Buchatskaya, C. Doersch, B. Avila Pires, Z. G. andM.G. Azar B. Piot K. Kavukcuoglu, R. Munos, and M. Valko, "Bootstrap your own latent-a new approach to self-supervised learning," in *Advances in Neural Information Processing Systems (NeurIPS 2020)*, vol. 33, Virtual, December 2020, pp. 21 271–21 284.

[23] X. Chen and K. He, "Exploring simple siamese representation learning," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Nashville, TN, USA, June 2021, pp. 15 745–15 753.

[24] S. Warnakulasuriya, G. Sutherland, and C.Scully, "Tobacco, oral cancer, and treatment of dependence," *Oral oncology*, vol. 41, no. 3, pp. 244–260, March 2005.

[25] K. Caughlin, R. Cuenca Martinez, G. P. Tortorelli, K. E. Higgins, R. Faram, J. A. Jo, and C. Busso, "Understanding bias in multispectral autofluorescence lifetime imaging: Are models sensitive to oral location?" in *IEEE Engineering in Medicine and Biology Society (EMBS 2024)*, Orlando, FL, USA, July 2024.